Q2. What is nominal encoding? Provide an example of how you would use it in a real-world scenario.

## Ans: What is Nominal Data?

Nominal data is “labeled” or “named” data which can be divided into various groups that do not overlap. Data is not measured or evaluated in this case; it is just assigned to multiple groups. These groups are unique and have no common elements.

EXAMPLE

 For a travel agency looking to launch a travel plan purely for a sample of individuals, this is the most basic question: Who loves to travel more?

* Men – M
* Women – W

Q3. In what situations is nominal encoding preferred over one-hot encoding? Provide a practical example.

1. **One Hot Encoding Technique:**In this technique, mapping is done for the different categories present in the feature to a vector consisting of 1 or 0 depending on the presence or the absence of the feature. The count for the number of vectors depends on the number of categories of data present in that particular feature. If for a particular feature, the number of categories present is huge then this technique increases the number of columns present in the dataset and reduces the learning rate of the algorithm significantly. The One hot encoding technique is usually applied to nominal data present in the data set. The application of One Hot Encoding for a particular column in a data set can be done as under.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Color | Color\_Blue | Color\_Red | Color\_Orange | Color\_Green | Color\_white |
| Blue | 1 | 0 | 0 | 0 | 0 |
| Red | 0 | 1 | 0 | 0 | 0 |
| Orange | 0 | 0 | 1 | 0 | 0 |
| Green | 0 | 0 | 0 | 1 | 0 |
| White | 0 | 0 | 0 | 0 | 1 |

Q4. Suppose you have a dataset containing categorical data with 5 unique values. Which encoding

technique would you use to transform this data into a format suitable for machine learning algorithms?

Explain why you made this choice.

Ans : I will use categorical technique because Categorical variables are usually represented as ‘strings’ or ‘categories’ and are finite in number and

the variables only have definite possible values. Further, we can see there are two kinds of categorical data-

* **Ordinal Data:** The categories have an inherent order
* **Nominal Data:** The categories do not have an inherent order